ASSESSED INTELLIGENCE

ASSESSED SIGNAL

July 2026

At the Intersection of Policy, Use, & Threat Intelligence

Illinois enacted the first state law in the country requiring independent third-party audits of frontier AI models, moving oversight from developer self-attestation to external verification and giving covered organizations a firm compliance date of January 1, 2027. The FBI warned that Russian intelligence services are now stealing Signal backup recovery keys to reconstruct target accounts, confirming that attackers have shifted past the application to the recovery layer most security programs do not monitor. Google's threat researchers documented the first case of an AI model being used to weaponize a vulnerability and produce working malware, while Anthropic found that the share of tracked malicious actors rated medium risk or higher rose from 33 percent to 56 percent in a single year. Current research shows that most generative AI pilots still fail to reach production and that most leaders cannot demonstrate they would pass an AI governance audit, exposing a widening distance between AI adoption and AI assurance. Organizations treating governance as verifiable evidence are positioned to deploy AI at scale, while others remain stalled by the risks they cannot measure.

Secure & Responsible Technology.

AssessedIntelligence.com

Illinois Made Independent Audits the Price of Deploying Frontier AI

On July 6, 2026, Governor JB Pritzker signed the Artificial Intelligence Safety Measures Act, Senate Bill 315, making Illinois the first state to require regular independent third-party safety audits of covered frontier AI systems.[1] The law takes effect January 1, 2027.[1] State frontier regulation until this point relied on developer self-attestation, meaning a company documented its own safety practices and asked regulators and customers to accept them. Illinois replaced that model with external verification, and that change reorders how covered organizations must prove their AI is safe. The Act applies to frontier developers, defined as companies with annual gross revenue above 500 million dollars that train models using computing power greater than 10 to the 26th operations, a scope that reaches roughly a dozen frontier labs.[2] [3] Covered developers must publish an annual frontier AI safety framework and pre-deployment transparency reports, commission annual independent audits conducted by auditors without financial conflicts of interest, and maintain confidential internal reporting channels with whistleblower protections.[1] [2] They must report a critical safety incident to state authorities within 72 hours, and within 24 hours when the incident poses an imminent risk of death or serious physical injury.[3] The Attorney General enforces the law, with civil penalties up to 1 million dollars for a first violation and 3 million dollars for subsequent violations.[2] [3] Although the Act takes effect January 1, 2027, the independent audit obligation phases in beginning in 2028, which gives covered developers a defined runway to stand up the required controls.[3] Illinois joins California, whose SB 53 became the first US frontier AI law in 2025, and New York, whose RAISE Act takes effect January 1, 2027; Illinois goes further than both by mandating independent audits rather than self-certification.[2] [4] [5]

This state activity now collides with a federal effort to centralize AI policy. A December 11, 2025 executive order directs the Attorney General to establish an AI Litigation Task Force to challenge state AI laws viewed as inconsistent with federal policy, and it threatens to make states with "onerous" AI laws ineligible for certain broadband funds.[6] A proposed federal moratorium on state AI laws was left out of the fiscal year 2026 defense authorization bill.[7] Organizations operating nationally therefore face divergent state obligations at the same time federal preemption remains unsettled.

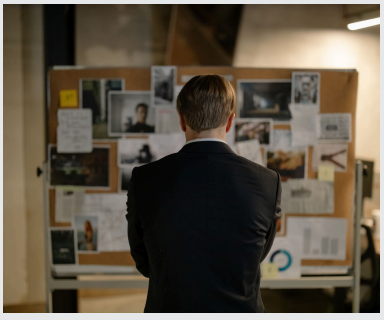
Industry leadership has named that uncertainty directly. On July 9, 2026, Fortune reported that Microsoft Vice Chair and President Brad Smith criticized the federal approach as "regulation without transparent or complete rules," warning that "without rules, businesses can't plan."[8] The strategic implication is specific: regulatory ambiguity is itself a compliance risk, and organizations should build to the most demanding applicable standard rather than the most lenient. The Illinois audit mandate connects directly to the ARISE Framework™ GOVERN and VALIDATE domains, because independent verification of AI safety claims is precisely what the VALIDATE domain formalizes. Organizations should treat mandated independent audit as the emerging baseline, not a jurisdictional exception.

1. Office of Governor JB Pritzker. Gov. Pritzker Signs Nation-Leading Artificial Intelligence Safety Law. July 6, 2026. https://gov-pritzker-newsroom.prezly.com/gov-pritzker-signs-nation-leading-artificial-intelligence-safety-law
2. McDonald Hopkins. Illinois Artificial Intelligence Safety Measures Act. July 2026. https://www.mcdonaldhopkins.com/insights/news/illinois-artificial-intelligence-safety-measures-act
3. Akerman LLP. Illinois SB 315: A State Strategy for Enduring National AI Safety Standards. 2026. https://www.akerman.com/en/perspectives/illinois-sb-315-a-state-strategy-for-enduring-national-ai-safety-standards.html
4. StateScoop. Illinois AI safety law requires audits of frontier models. July 2026. https://statescoop.com/illinois-ai-safety-law-audits-frontier-models/
5. Wiley. New York Finalizes RAISE Act for Frontier AI Models; Law Takes Effect January 1, 2027. 2026. https://www.wiley.law/alert-New-York-Finalizes-RAISE-Act-for-Frontier-AI-Models-Law-Takes-Effect-January-1-2027
6. The White House. Executive Order on a National Policy Framework for Artificial Intelligence. December 11, 2025. https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/
7. StateScoop. State AI law moratorium omitted from 2026 defense bill as Trump signals executive order. December 2025. https://statescoop.com/state-ai-law-moratorium-omitted-2026-defense-bill-trump-eo/
8. Fortune. Microsoft's Brad Smith on Washington's AI policy: regulation without transparent rules. July 9, 2026. https://fortune.com/2026/07/09/microsoft-brad-smith-washingtons-ai-policy-regulation-without-transparent-rules/

Russian Intelligence Moved the Attack to the Recovery Key

On June 26, 2026, the FBI Internet Crime Complaint Center issued public service announcement I-062626-PSA warning that Russian intelligence services are targeting Signal messenger backup recovery keys.[1] The finding matters because it describes a deliberate move past the encrypted application to the recovery layer beneath it. Signal's message encryption remains sound; the attackers do not break it, they obtain the credential that restores an account somewhere else.

The activity is attributed to two clusters that Google Threat Intelligence Group and Mandiant track as UNC5792 and UNC4221.[1] The actors impersonate Signal support and automated accounts, then use fabricated pretexts such as a mandatory two-factor prompt or a sync error to induce a target into revealing the 64-character backup recovery key.[1] With that key, an actor restores the victim's backup onto attacker-controlled hardware and reads historical direct and group messages.[1] A critical detail for response planning is that creating a new Signal account with the same phone number does not invalidate a stolen key; a victim must generate a new backup recovery key, and even that cannot undo access to backups already taken.[2] The confirmed targets include current and former government officials, military personnel,



journalists, and Ukrainian officials.[1] The technique is notable for what it does not require: no malware on the device, no exploit against the platform, and no interception of live traffic. It requires only that a trusted person be persuaded to share one credential, which places the failure at the identity and awareness layer rather than the technical one.

This campaign sits inside a broader shift in how attackers use AI. On June 3, 2026, Anthropic published an analysis of 832 accounts banned for malicious cyber activity between March 2025 and March 2026, and found that roughly 67 percent of those actors used AI to write malware.[3] The more consequential figure is the trajectory of severity: the share of tracked actors classified as medium risk or higher rose from 33 percent in the first half of the study to 56 percent in the second.[3] AI is compressing the skill curve that once separated capable attackers from opportunistic ones.

The evidence points to a specific structural conclusion. The defensible perimeter is no longer the endpoint or the application; it is identity, recovery, and configuration state. Detection programs that watch for malware execution but not for credential recovery and account restoration events are measuring the wrong surface. This maps to the ARISE Framework™ DETECT and RESPOND domains. Organizations should extend monitoring to recovery and linked-device flows, and their incident response playbooks must treat a stolen recovery key as a full account compromise, because that is what it is. The same logic applies well beyond one messaging application; any system that permits backup, restore, or device linking presents the same recovery-layer exposure, and security programs should assess each of them on that basis rather than assuming end-to-end encryption resolves the risk.

1. FBI Internet Crime Complaint Center. Russian Intelligence Services Targeting Signal Messenger Backup Recovery Keys. PSA I-062626-PSA. June 26, 2026. https://www.ic3.gov/PSA/2026/PSA260626
2. BleepingComputer. FBI: Russian hackers now target Signal backup recovery keys. June 26, 2026. https://www.bleepingcomputer.com/news/security/fbi-russian-hackers-now-target-signal-backup-recovery-keys/
3. Anthropic. Mapping a year of AI-enabled cyber threats using the MITRE ATT&CK framework. June 3, 2026. https://www.anthropic.com/news/AI-enabled-cyber-threats-mitre-attack

AI Weaponized a Vulnerability and Wrote the Malware

Most organizations have deployed AI faster than they can govern it, and the current research quantifies the gap. MIT's NANDA initiative reported in 2025 that roughly 95 percent of enterprise generative AI pilots fail to reach measurable profit-and-loss impact.[1] Gartner projects that more than 40 percent of agentic AI projects will be canceled by the end of 2027 because of unclear business value, rising costs, or inadequate risk controls.[2] Deployment is not the constraint; the constraint is the assurance required to move from pilot to production safely.

The governance data is more direct. A Grant Thornton survey released in April 2026 found that 78 percent of senior US business leaders are not fully confident they could pass an independent AI governance audit within 90 days, and only 12 percent describe their workforce as truly AI-ready.[3] The security consequences of that gap are already priced. IBM's 2025 Cost of a Data Breach report found that breaches involving shadow AI averaged 4.63 million dollars, that organizations with high levels of shadow AI paid roughly 670,000 dollars more per breach, and that 97 percent of AI-related breaches occurred in organizations lacking proper AI access controls.[4] Varonis reported in 2025 that 99 percent of organizations have sensitive data that AI tools can readily surface.[5] A Cloud Security Alliance survey conducted with Token Security and released in April 2026 found that 82 percent of enterprises have unknown or unmanaged AI agents operating in their environments, and that 65 percent had experienced at least one security incident caused by an AI agent in the prior year.[6]

The inference is straightforward. The pilot-to-production gap is a governance gap, and an unmanaged agent is an unmanaged risk that carries the same access as the identity it runs under. Assessed Intelligence research finds that 76 percent of critical-sector companies have AI risk governance gaps, which is consistent with every external dataset above. This condition maps to the ARISE Framework™ GOVERN, MANAGE, and IDENTIFY domains. Organizations must inventory the AI agents already operating inside their environments, enforce access controls scoped to those agents, and establish the evidence trail an audit will require, because the audits described elsewhere in this edition are becoming mandatory. The organizations that close this gap first will not simply reduce risk; they will convert governance into a deployment advantage, because the capacity to prove an AI system is controlled is what allows it to move into production at all.

1. Fortune. MIT report: 95% of generative AI pilots at companies are failing. August 18, 2025. https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo/
2. Gartner. Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027. June 25, 2025. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027
3. Grant Thornton. Survey on the AI proof gap. April 13, 2026. https://www.grantthornton.com/insights/press-releases/2026/april/grant-thornton-survey-on-ai-proof-gap
4. IBM. Cost of a Data Breach Report 2025. July 2025. https://www.ibm.com/reports/data-breach
5. Varonis. 2025 State of Data Security Report. June 2025. https://www.varonis.com/blog/state-of-data-security-report
6. Cloud Security Alliance. New Survey Reveals 82% of Enterprises Have Unknown AI Agents in Their Environments. April 21, 2026. https://cloudsecurityalliance.org/press-releases/2026/04/21/new-cloud-security-alliance-survey-reveals-82-of-enterprises-have-unknown-ai-agents-in-their-environments

ISO 42001 Is Becoming the Evidence Standard for AI Management

As independent audits move from optional to mandatory and enterprise buyers demand proof of AI oversight, organizations need a management system standard purpose-built for AI. ISO/IEC 42001 is that standard: the first international AI management system standard, published in 2023, defining requirements for AI policy, risk management, data governance, system lifecycle controls, human oversight, and continual improvement.[1] It exists because information security controls alone do not address AI-specific risks such as model drift, training-data governance, and unsafe autonomous behavior.

Adoption is early but accelerating. Roughly 350 organizations held ISO 42001 certificates by the spring of 2026, a figure assembled from certification-body and company announcements rather than a single official register, so it should be read as an informed estimate.[2] Boston Consulting Group announced on January 27, 2026 that it was among the first 100 organizations globally to be certified.[3] The standard is designed to pair with ISO/IEC 27001, the information security management standard whose transition to the 2022 revision closed on October 31, 2025.[4] The two share a harmonized management system structure, which lets an organization run a single integrated governance program: ISO 27001 protects information assets, and ISO 42001 adds the AI management system on the same audit and risk machinery.[1] [4]

This is where the recurring theme of this edition resolves. When frontier regulation requires independent verification, when threat actors weaponize models and unmanaged agents, and when most organizations cannot evidence their AI oversight, technology governance becomes the operative control surface. Governance is the new perimeter, and a certifiable management system is how an organization demonstrates that the perimeter holds.

Certification is therefore shifting from a differentiator to a baseline expectation embedded in procurement questionnaires and regulatory scrutiny. External standards define the destination; they do not, by themselves, produce continuous evidence between annual audit cycles. The ARISE Framework™ provides that continuous assurance layer across its seven domains, GOVERN, MANAGE, IDENTIFY, PROTECT, DETECT, RESPOND, and VALIDATE, mapping the evidence an organization generates every day to the standards it will be measured against. Organizations should pursue ISO 42001 certification as the external proof point and operate a continuous assurance program beneath it, because an audit certifies a moment while risk accrues continuously. A certificate confirms that a management system met the criteria on the day it was examined; it does not attest to the model retrained last week or the agent granted new permissions yesterday. Closing that interval between certification and reality is the work that separates organizations that can defend their AI when it matters most from those that can only describe it

1. International Organization for Standardization. ISO/IEC 42001: what it is and why it matters. 2024. https://www.iso.org/home/insights-news/resources/iso-42001-explained-what-it-is.html
2. Atoro. How many companies are ISO 42001 certified? July 2, 2026. https://atoro.io/how-many-companies-are-iso-42001-certified/
3. Boston Consulting Group. BCG Certified to International Standard for AI Management Systems. January 27, 2026. https://www.bcg.com/news/27january2026-bcg-certified-international-standard-ai-management-systems
4. LRQA. Preparing for the ISO/IEC 27001:2022 transition by October 2025. 2025. https://www.lrqa.com/en-us/insights/articles/preparing-for-iso-270012022-transition-by-october-2025/

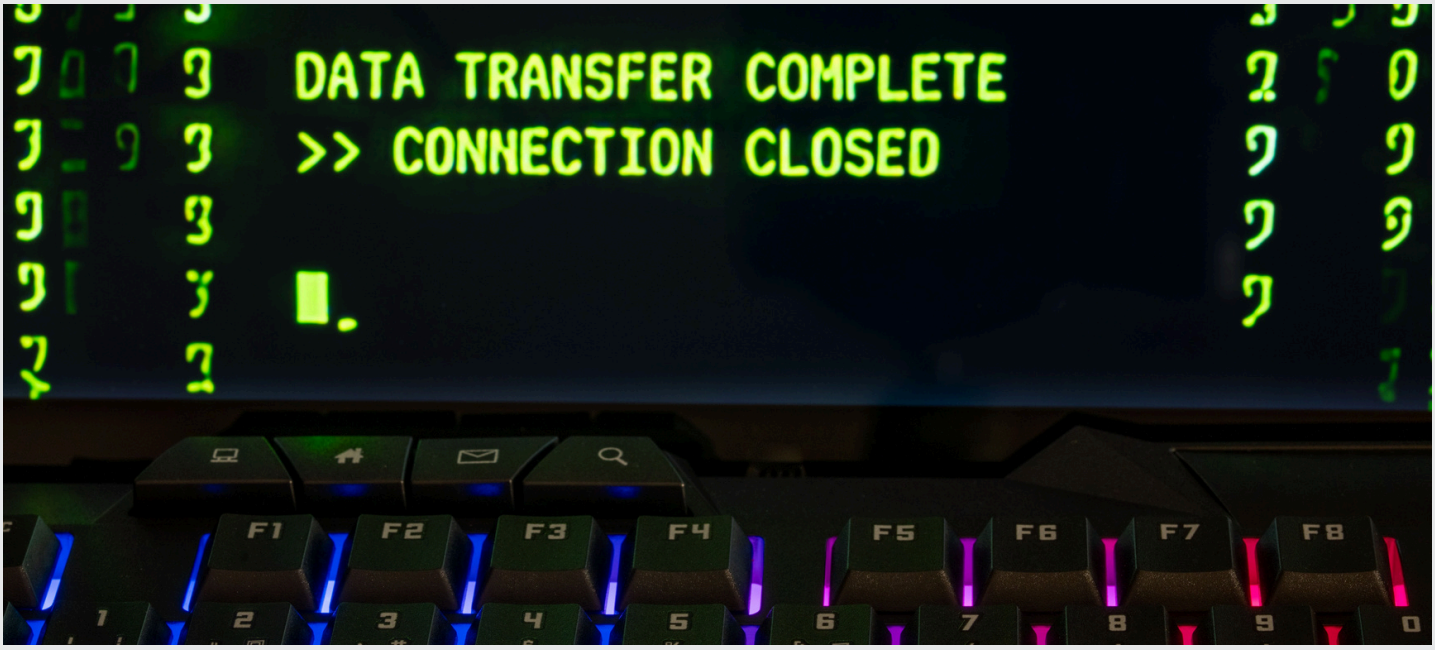
State-Linked Actors Ran Influence Operations on Commercial AI Models

In its June 2026 threat report, OpenAI disclosed that it disrupted two clusters of ChatGPT accounts it linked to actors originating in China, used to generate content for influence operations aimed at United States audiences.[1] The significance is that the same generative capability organizations adopt for productivity is being used at scale to manufacture political narratives.

The report describes two campaigns. One, which OpenAI tracked as "Data Center Bandwagon," pushed US-targeted narratives claiming that AI data centers raise local electricity costs. The other, "Tech and Tariffs," produced anti-tariff cartoons, with operator instructions to exclude imagery of specific Chinese leadership.[1] OpenAI associated one cluster with individuals tied to a private Chinese technology company serving provincial-government clients, and another with individuals associated with China's public security apparatus.[1] This activity is content generation for information operations, not automated intrusion, and organizations should hold that distinction rather than collapse every AI misuse story into a single category.

The consequential shift is cost. Influence operations historically required staff to write, translate, and localize material; a model collapses that labor. The output is not necessarily more sophisticated than human-produced propaganda, but there is far more of it, produced faster and adapted to more audiences. This is the same economic pattern Anthropic documented across malicious cyber actors, where the share rated medium risk or higher rose from 33 percent to 56 percent in a year as AI lowered the effort required to operate.[2] Detection that relies on volume and linguistic fingerprints becomes less reliable as the generation source improves.

For organizations, the exposure is twofold. Brand and market narratives can be shaped by synthetic content at scale, and internal staff who consume that content are a target for the narratives it carries. This maps to the ARISE Framework™ IDENTIFY and DETECT domains. Communications, security, and risk functions should coordinate monitoring of synthetic narrative activity that touches their sector, and they should establish an assessed process for verifying the provenance of external content that informs decisions, rather than treating influence operations as a purely governmental concern.



1. OpenAI. June 2026 Threat Report. June 2026. https://cdn.openai.com/pdf/96b559fa-c165-4575-805d-e636909e2f78/June-2026-Threat-Report.pdf
2. Anthropic. Mapping a year of AI-enabled cyber threats using the MITRE ATT&CK framework. June 3, 2026. https://www.anthropic.com/news/AI-enabled-cyber-threats-mitre-attack

ASSESSED INTELLIGENCE

Secure & Responsible Technology.

YOU'RE BUILDING THE FUTURE. DON'T LET EMERGING RISK STALL YOUR MOMENTUM.

SALES@ASSESSEDINTELLIGENCE.COM
EMEASALES@ASSESSEDINTELLIGENCE.COM